DOCUMENT RESUME

ED 079 381 TM 002 981

AUTHOR Rimoldi, H. J. A.
TITLE Evaluation and Training of Clinical Diagnostic Skills. Publication No. 41.
INSTITUTION Loyola Univ., Chicago, Ill. Psychometric Lab.
SPONS AGENCY Commonwealth Fund, New York, N.Y.
REPORT NO Pub-41
PUB DATE 63
NOTE 11p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Clinical Diagnosis; Medical Case Histories; *Medical Education; Medical Students; Performance Criteria; *Performance Tests; Professional Training; Scoring; Statistical Analysis; *Student Evaluation; Technical Reports
IDENTIFIERS *Test of Diagnostic Skills

ABSTRACT

The Test of Diagnostic Skills is intended to explore the way a physician or a medical student solves a clinical problem through analysis of the type and sequence of the questions he/she asks. The emphasis is on processes used to reach a conclusion rather than on the accuracy of the conclusion itself. Information based on a real clinical case is written on removable cards placed in overlapping flat pockets. The questions that may be asked are written on the top edges of the cards. The information pertinent to a question is contained on the reverse side of the card. After reading preliminary information about the case, the subject is requested to reach a diagnosis by asking as many questions as he wishes, in any order, from the questions presented to him. The subject is instructed to read the information contained on the reverse side of the chosen card before asking the next question. The questions asked, and those not asked, are recorded. The test usually consists of 50 to 80 cards. Several scoring methods have been developed to study the performance of junior and senior medical students and physicians; the results reported were obtained with a sample of approximately 90 juniors, 130 seniors, and 40 physicians. The mean number of questions asked by each group for the three parts of the test, clinical interview, physical examination and laboratory data, was caluclated; it was found that juniors and physicians vary more than seniors in the number of questions asked. Also computed were utility indexes for each question, the performance curve of each subject, and values used to score students in terms of the physicians' performance. (For related document, see TM 002 982.) (KM)

FILMED FROM BEST AVAILABLE COPY

ED 079381

A L
TM

PSYCHOMETRIC LABORATORY

EVALUATION AND TRAINING OF CLINICAL DIAGNOSTIC SKILLS

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF
EDUCATION POSITION OR POLICY

TM 002 981

by

H. J. A. Rimoldi

Loyola University
Chicago, Illinois

1963

Publication No. 41

EVALUATION AND TRAINING OF CLINICAL DIAGNOSTIC SKILLS*

by

H. J. A. Rimoldi, M.D., Ph.D.

While the problem of evaluating the students' knowledge of medical information is relatively simple, this can not be said in relation to the evaluation of their clinical diagnostic skills. Diagnostic skill implies more than the mere accumulation of factual knowledge. Besides personality characteristics, and the interaction doctor-patient that we shall not discuss, there is, in the diagnostic situation, a selective search for information, a test of hypotheses that should become more clearly defined as the process develops, a verification of previously held hunches and hypotheses and a direction towards a goal whose characteristics should emerge clearly provided a diagnosis is reached. As a matter of fact, the diagnostic process is a dynamic and plastic situation that changes continuously as a function of what has preceeded, of the recently acquired information and of the subgoals, that at the time, the diagnostician is following.

This attempt at formulating some of the variables that may be experimentally controlled in order to study the clinical diagnostic process indicates that diagnosing is more than remembering isolated facts. It brings into the picture analysis and synthesis of information, to fit the specific case at specific moments during the diagnostic process. Whether this is only a matter of knowledge and/or training, and whether each identifiable step can be thoroughly understood is a problem yet to be solved.

An instrument to appraise clinical diagnostic skills should permit the study of the features previously described.

"The Test of Diagnostic Skills aims at exploring how a physician or a medical student solves a clinical problem by analyzing the type and the sequence of questions that he asks" (Rimoldi, Haley, and Fogliatto, 1962). It is a special application of a technique that has been used by myself and associates in order to explore a good number of psychological problems. (Rimoldi, 1955; Tabor, 1959; Haley 1960; Mohrbacher, 1960; Rimoldi, 1960; Gunn, 1961; Rimoldi, 1961; Rimoldi and Devane, 1961; Rimoldi, Fogliatto, Haley, Reyes, Erdmann and Zacharia, 1962; Rimoldi, Meyer, Meyer and Fogliatto, 1962). Its basic rationale departs from usual testing procedures. It aims at making explicit the sequence of questions that a subject asks when solving a problem, that is, his "tactic". The answers given to each question are a fixed property of the test while the questions are either generated by the

---

* This is a summary of research performed during the tenure of a grant from the Commonwealth Fund of New York; Principal Investigators: H. J. A. Rimoldi and John T. Cowles.

subject (Rimoldi, Fogliatto, Haley, Reyes, Erdmann, and Zacharia, 1962) or chosen, among a large number of possible questions. We are not concerned with selection of the right answer in a multiple choice situation, or with categorizing an answer as right or wrong, ignoring how the subject reached it. As a matter of fact, it has been demonstrated (Rimoldi and Devane, 1961; Rimoldi, Fogliatto, Haley, Reyes, Erdmann, and Zacharia, 1962; Rimoldi, Haley, and Fogliatto, 1962) that the same answer may result from different processes, that mirror more clearly individual differences in performance, than final answers do.

The experimenter knows the number of questions asked, type of questions, their order in the sequnce, and any comments, written or verbal, that the subject may wish to advance. It should be clear that the operations that can be performed with this type of data should consider the dimension "order of questions", and that a satisfactory evaluation of the results can not be performed using some of the traditional techniques employed in dealing with most of the known psychological tests.

The actual test is based on a real clinical case transcribed into a set of cards. "Artificial" cases to fit specific purposes may be used, but our experience in this area has been discouraging. The information is written on removable cards contained in flat pockets which partially overlap and are evenly arranged on a display folder. "On the top edge of the numbered cards", "the questions that the examiner may ask are indicated" (Rimoldi, Haley, and Fogliatto, 1962). Drawing a card and looking at the reverse side, the pertinent information is obtained. The cards contain questions that the doctor may want to ask verbally, or may refer to manipulations that he wishes to perform, laboratory tests that he may want to order, and so forth.

The subject is presented first with information usually available from the hospital admission chart, patient's complaints and other aspects of his clinical history. After reading this, he is requested to reach a diagnosis of the case by asking as many questions as he wishes from those presented to him in any order he wants. He is instructed to read the information contained on the reverse side of the chosen card before asking the next question. The subject is free to stop drawing cards at any desired time. The experimenter, or the subject, records the question asked, as well as those not asked. Usually these tests consist of 50 to 80 cards. The technique is adaptable to a large number of situations. The experimenter can control the set of questions presented to the subjects as well as the information provided including its mode of presentation, i.e., photographs, actual E.K.G. records, verbal descriptions, etc.

It should be noticed that in the real diagnostic situation the doctor himself generates the questions he wants to ask. Our experience has shown that it is possible to develop a set of questions that in all likelihood will cover all those that the doctor may want to ask. We are also aware

that the presentation of the cards may suggest questions that otherwise may not have been asked. The technique here described can not claim perfect validity, as no known assessment instrument does. But if medical education develops and improves diagnostic ability, changes in performance in the Test of Diagnostic Skills should be related to changes in medical training and should reflect what previous studies and experience have demonstrated to occur.

Several scoring methods have been developed and employed to study the performance of junior and senior medical students and physicians from several medical schools.

The mean number of questions, "cards", asked by juniors, seniors and physicians in the whole test and in specific areas referring to clinical interview (Part I), physical examination (Part II) and laboratory data (Part III) shows that: juniors ask more questions than seniors and these more than physicians in the whole test, in Part I and in Part II. In Part III, seniors ask less questions than physicians and than juniors. These differences are greater in Part I of the test (clinical interview). These findings (Rimoldi, Haley, and Fogliatto, 1962) have been tentatively interpreted as indicating that with increased medical training an important change in diagnostic ability relates to the interpretation of interview data, due to increased ability to use information that patients sometimes present in obscure ways.

Juniors and physicians vary more than seniors in the number of questions asked. In the light of other evidence (content of the questions asked) this may indicate that training eliminates individual differences, so that seniors are more homogeneous in the number of questions they request. "With further training the individual differences will tend to reappear but now they will be more closely related to the nature of the case under study" (Rimoldi, Haley, and Fogliatto, 1962). It may be worthwhile to consider the possibility of accelerating training in diagnosis early in the medical studies to reach as soon as possible a certain base line from which to build for further improvement.

The same type of results was obtained when the same subjects were examined in two successive years. At the senior level the number of questions asked was less, the difference being greater for Part I of the test. The analysis of variance of these results indicated the existence of a significant interaction between administration of the tests and subjects, that is, the training period between the first and second administration, though reducing the number of questions asked, affected each subject differentially. (Haley, 1960). A word of caution is now in order: the meaning of the score number of questions may have other interpretations than the one here suggested and the thoroughness with which a student interprets information may only be partially related to this score.

Utility indexes for each question are defined as the ratio between the number of times that a question is selected by the members of a group and the total number of members in the group. It follows that questions may have utility indexes from zero to 1.00 and that these may vary with each group. It can be assumed that a "popular" question is perceived as more useful than one that is less popular, hence the use of the expression "utility index". The values of the utility indexes of some questions vary greatly from group to group, i. e., questions seldom asked by juniors are preferred by physicians and vice versa. The content analysis of these questions shows some of the pitfalls that occur early during medical training. (Rimoldi, Devane, and Grib, 1958a, 1958b). Nevertheless, utility indexes for the same cards as obtained in four different groups of physicians, showed similar values (Devane, Rimoldi, and Haley, 1959). This may indicate a high similarity in perceiving the usefulness of the questions in the test, and makes it permissible to pool together the performance of these four groups for developing a set of more stable utility indexes.

The performance curve of each subject can be graphically presented by adding the utility indexes of the questions asked by a subject in the order in which they were selected. Subjects can be scored using utility indexes developed either from his own group, from the physicians' group or in any other logical and permissible fashion. If we are more interested in evaluating a student's performance in terms of a prescribed aim rather than in terms of its standing in relation to his group (as evaluations are often performed) then the utility indexes to be used are those obtained from the physicians, under the reasonable assumption that this group is more experienced and knows more about diagnosis than the students do. Though we have experimented with several approaches, here we shall use only results obtained using utility indexes based on physicians' performance in the Test of Diagnostic Skills.

Assume that the most efficient performance in the test corresponds to a sequence of questions that maximizes the sum of the utility indexes at each step. Ordering the questions from high to low utility indexes, cumulating them and graphing the successive values, a maximum curve is obtained (Table I and Figure I). The minimum curve results from ordering the questions in reverse order. Between maximum and minimum curve we have thus far always obtained ellipsoids, but these will degenerate into a straight line when all the utility indexes have the same value, that is, when the questions do not have differentiating power. In the case of Figure I the sume of the utility indexes is 4.22. Then 4.22 cards with utility indexes of 1.00 should reach the same height as the maximum curve. Since the test has 11 cards (questions) then 6.78 of them should have 0 utility index in the condition of maximum differentiating power. (Rimoldi, Devane, and Haley, 1961). Plotting these values (Figure I) a parallelogram is obtained. The ratio between the area of the ellipsoid and the parallelogram will be greater whenever the ellipsoid is nearer in size to the parallelogram. These ratios were always found to be greater for physicians than for seniors and for seniors greater than for juniors. When examined in

relation to the difficulty of the test, the ratios are always greater the easier the test. Confirmatory results were obtained when comparing the performance of the same group of subjects after one year of increased medical experience. (Haley, 1960).

The slope of the maximum curve* can be interpreted to indicate that the information that can be gained at every step in the process is a constant ratio of the information yet to be gained. This value increases from juniors to seniors and is highest for physicians. When different tests are used with the same group of subjects it is found that the slope of the curve is higher for the easier diagnostic problems than for the more difficult ones. Thus in the former the maximum curve is steeper than in the latter. Confirmatory evidence of this was found when comparing the same group after one year of medical training (Haley, 1960). This indicated what was logically expected: physicians obtain more information at every step than either junior or senior medical students.

The performance of each student can be graphically presented (Figure I) by accumulating the utility indexes of the cards successively selected. In Figure I, subject j performed "better" than k. Analysis of these curves, in general and at each step is of value in interpreting how the subject proceeds. Plateaux tend to disappear as training increases (Haley, 1960).

The method described does not consider order and makes rather strong assumptions concerning the best type of performance. But a question may have different utility according to its position in the sequence. By counting the number of times that each question is selected in any possible order as well as the number of times that it was not selected (0 order) and expressing this as a ratio of the total number of possible selections, a new set of values is obtained. If a problem is administered to 10 subjects and has 5 possible questions, then the total number of possible selections will be 50.

These values can be used to score students in terms of the physicians' performance. Each question asked will receive a value depending on its position in the sequence. This can be represented graphically by accumulating the successive values of the questions. It has been shown that the same set of questions asked in different orders gives different curves (Rimoldi, 1961; Rimoldi and Haley, 1962; Rimoldi, Haley, and Fogliatto, 1962). This corresponds to the assumption that the value of a question depends on its position in the sequence. These curves are important in making objective the subject's performance. Each step can be related to each question and the study of plateaux, redundancy, following of cues, verifying hypotheses and so forth, can be discussed with the subject to whom the test was administered. (Rimoldi, 1962). A set of problems of graded difficulty or built to test specific points can be prepared. After administering them to the students it would be an easy matter for the instructor to discuss their performance curves and thus gain information as to their actual diagnostic ability (as appraised by the test) and suggest ways for its improvement,

* The formula that fits our observation is of the form $y=c(1-e^{-bx})$. $b>0$, where y = sum of utility indexes, x = number of questions asked, b = slope and c = asymptote.

Fig. 2 Cumulative scores representing the final values of performance curves for test 2 of students scored in terms of surgeons' norms and clinicians' norms.

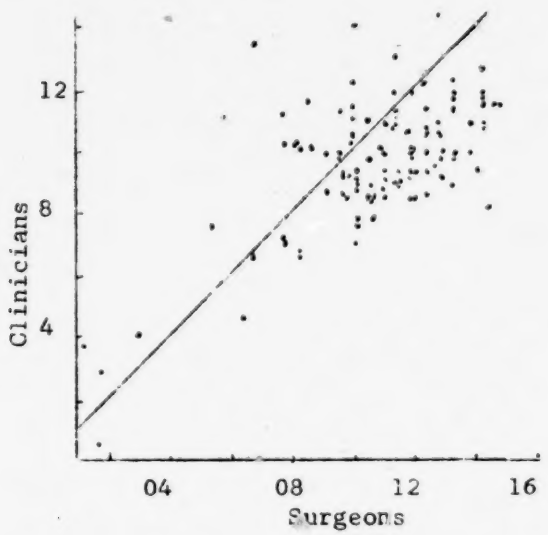
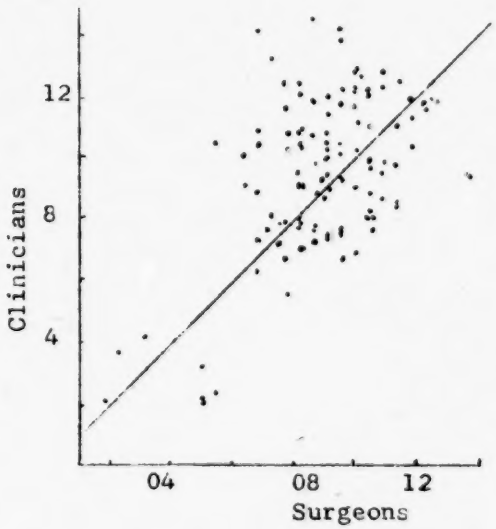


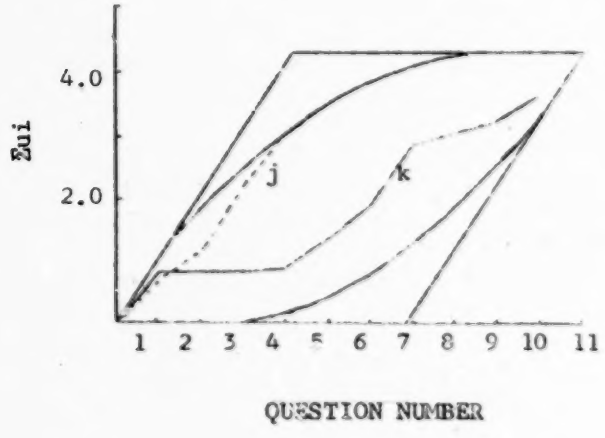
Fig. 3 Cumulative scores representing the final values of performance curves for test 4 of students scored in terms of surgeons' norms and clinicians' norms.



Data for Cumulation of Utility Indexes

Question Number	u_i	Σui maximum	Σui minimum
2	.98	.98	4.22
1	.80	1.78	3.24
8	.65	2.43	2.44
3	.50	2.93	1.79
7	.48	3.41	1.29
5	.37	3.78	.81
10	.20	3.98	.44
11	.15	4.13	.24
6	.08	4.21	.09
4	.01	4.22	.01
9	.00	4.22	.00

Fig. 1 Maximum and minimum curves for data of Table and performance curves for subjects j and k.



avoiding misuse of data, logical pitfalls, irrelevancy of cues, insufficient verification of information, etc. (Rimoldi, 1962; Haley, 1962).

Since a value for questions not asked is also available, each subject can be scored on these 0 order questions. They tell, as it were, the other side of the story, and at that, an interesting one. - Several recent technical developments, that will not be detailed, have been made (Rimoldi, 1961; Rimoldi, and Haley, 1962; Rimoldi, Haley, and Fogliatto, 1962; Rimoldi, Haley, Fogliatto, and Erdmann, 1962, unpublished). These aim at evaluating tactics in terms of different hypotheses, of their stability, their classification into families, etc. Others refer to ways in which problems can be controlled in relation to the complexity of their logical structures and their content (Rimoldi, Fogliatto, Haley, Reyes, Erdmann, and Zacharia, 1962).

An example of this type of assessment is presented in Figure 2 and Figure 3. Students were scored in terms of two sets of norms developed separately from the performance of clinicians and surgeons in two tests of diagnostic skills: Test 2 and Test 4. Each point in the graph represents a student and corresponds to the sum of the values of the question asked in terms of clinicians' (ordinate) and surgeons' norms (abscissa). Test 2 corresponds to a surgical case; Test 4 may not be considered to be predominantly surgical. Figures 2 and 3 show that most of the students fall on the surgeons' side in Test 2 and on the clinicians' side on Test 4.

The results reported in this article were obtained with a sample of spproximately 90 juniors, 130 seniors and 40 physicians. It can be said that the Test of Diagnostic Skills is sensitive to levels of training in diagnostic ability, that it permits the evaluation of some aspects of the diagnostic process, that specific types of performance can be studied and that it can be a useful teaching devise to accelerate and improve diagnostic ability. All this requires further confirmation. As an instrument for selection, problems in the basic sciences could be prepared, and performance curves developed for each candidate. On the whole, the type of evaluation described in this article makes explicit, at least partially, some of the facets of the thinking processes that are not so clearly analyzed with some of the usual testing procedures. Complementing usual evaluation procedures with those described in this study may improve the teaching and the training in medicine.

REFERENCES

Devane, J., Rimoldi, H.J.A., Haley, J. (1960). A comparison of the Performance of Two Student Groups and Physicians in the Tests of Diagnostic Skills. Loyola Psychometric Laboratory, Publication No. 10, Chicago, Illinois.

Gunn, H.E. (1961). Appraisal of Personality Parameters in Terms of Processes. (unpublished Ph. D. dissertation). Loyola University, Chicago, Illinois.

Haley, J.V. (1960). The effect on Learning on Performance in the Test of Diagnostic Skills. Loyola Psychometric Laboratory, Publication No. 11, Chicago, Illinois

Haley, J.V. (1962). Effects of Training on the Test of Diagnostic Skills. (paper presented at the First Annual Conference on Research in Medical Education, Association of American Medical Colleges, Los Angeles, California).

Mohrbacher, J. W. (1960). An Analysis of the Interdisciplinary Evaluation of Organic Pathology in a Child Guidance Setting. (unpublished Ph. D. dissertation). Loyola University, Chicago, Illinois.

Rimoldi, H.J.A. (1955). A Technique for the Study of Problem Solving. Educ. & Psych. Measurement, 15, 450.

Rimoldi, H.J.A. (1960). Problem Solving as a Process. Educ. & Psych. Measurement, 20, 449.

Rimoldi, H.J.A. (1961a). The Test of Diagnostic Skills. J. Med. Educ., 36, 73.

Rimoldi, H.J.A. (1961b). L'Étude des Processus Psychologiques. Travail Humain (France), 3-4, 225.

Rimoldi, H.J.A. (1963). Rationale and Application of the Test of Diagnostic Skills. J. Med. Educ., 38, 5, 364-368.

Rimoldi, H.J.A., Devane, J.R. (1961). Training in Problem Solving. Loyola Psychometric Laboratory, Publication No. 21, Chicago, Illinois.

Rimoldi, H.J.A., Devane, J.R., Grib, T.F. (1958a). Testing Skills in Medical Diagnosis. Loyola Psychometric Laboratory, Publication No. 2, Chicago, Illinois.

Rimoldi, H.J.A., Devane, J.R., Grib, T.F. (1958b). Supplementary Report on Testing Skills in Medical Diagnosis. Loyola Psychometric Laboratory, Publication No. 3, Chicago, Illinois.

Rimoldi, H.J.A., Devane, J.R., Haley, J.V. (1961). Characterization of Processes. Educ. & Psych. Measurement, 21, 383.

Rimoldi, H.J.A., Fogliatto, H.M., Haley, J.V., Reyes, I.O., Erdmann, J.B., Zacharia, R.M. (1962). Training in Problem Solving. Loyola Psychometric Laboratory, Publication No. 27, Chicago, Illinois.

Rimoldi, H.J.A., Haley, J.V. (1962). Sequential Evaluation of Problem Solving Processes. Loyola Psychometric Laboratory, Publication No. 22, Chicago, Illinois.

Rimoldi, H.J.A., Haley, J.V., Fogliatto, H.M. (1962). The Test of Diagnostic Skills. Loyola Psychometric Laboratory, Publication No. 25, Chicago, Illinois.

Rimoldi, H.J.A., Haley, J.V., Fogliatto, H.M., Erdmann, J.B., (1962). Program for the Study of Thinking. Loyola Psychometric Laboratory, Chicago, Illinois (unpublished).

Rimoldi, H.J.A., Meyer, R.A., Meyer, M.L., Fogliatto, H.M., (1962). Psychobiological Mechanisms in Complex Mental Processes and Their Changes With Age. Loyola Psychometric Laboratory, Publication No. 24, Chicago, Illinois.

Tabor, A. (1959). Process Analysis of Rorschach Interpretation. (unpublished Ph. D. dissertation). Loyola University, Chicago, Illinois